

Analyzing the Role of Spinal Joint Dynamics in the Movement of a Sprawling Robot

Merve Atasever^{1,*}, Azhang Nazaripouya^{2,*}, Ali Okhovat², John Nisbet², Omer Kurkutlu^{2,3},
Jyotirmoy V. Deshmukh¹, Yasemin Ozkan Aydin²

Abstract—Sprawling locomotion in vertebrates, particularly salamanders, illustrates how body undulation and spinal mobility can support stable and efficient movement. Translating these principles to small robotic platforms remains challenging because increasing morphological flexibility also increases the control burden. In this work, we study a simplified salamander-inspired quadruped with four 2-DoF legs and a single active spinal joint. Rather than learning all joint motions end-to-end, we evaluate a hybrid controller in which a Hildebrand-style gait provides structured limb coordination while reinforcement learning controls only the spinal joint. This design reduces the policy search space and improves practical deployability, but also restricts gait flexibility compared with fully learned locomotion policies. Through simulation and hardware experiments, we compare fixed-spine, fully learned, torque-limited, and hybrid spine-actuated controllers on target-directed crawling tasks. Our results show that active spinal control improves forward displacement and rough-terrain traversal relative to the fixed-spine Hildebrand baseline, while avoiding some high-frequency behaviors observed in unconstrained learned policies. These findings suggest that combining biologically inspired gait priors with low-dimensional learned spinal actuation is a promising but deliberately constrained approach for practical sprawling-robot locomotion.

I. INTRODUCTION

Among the vertebrates, salamanders, with their unique ability to transition between walking and swimming gaits, highlight the role of spinal mobility in locomotion [1]. A flexible spine enables undulation of the body, a wavelike motion along the spine that supports diverse movements, aiding navigation over uneven terrains and obstacles [2]. Such biomechanical principles have been the inspiration for the development of robotic systems that are capable of mimicking these natural movements and can overcome the limitations of rigid structures, achieving greater efficiency and adaptability in complex environments [1], [3], [4].

In both animals and robots, the coordination between body undulation and limb movement is of high significance for effective locomotion [4]–[8]. For salamanders, this coordination is essential for creating propulsion



Fig. 1. **Biological and robotic model.** (left) The fire salamander (*Salamandra salamandra*) is a common species of salamander found in Europe. (middle) The robot model in the Mujoco simulation world. (right) 3D printed robotic salamander.

during swimming and ensuring stability while walking [5]. Replicating this morphology in robotics presents a significant challenge because effective sprawling locomotion requires coordinated control of limb motion, body bending, contact timing, and ground interaction across multiple degrees of freedom [5], [9], [10].

Yet environmental uncertainties, such as surface irregularities and variations in friction, can significantly disrupt body-limb coordination and cause discrepancies between predictions from mathematical models and real-world outcomes. Addressing this challenge requires the development of sophisticated control strategies capable of dynamically adapting to uncertain conditions while maintaining efficient locomotion. Building on biomechanical insights, incorporating Deep Reinforcement Learning (DRL) techniques with sensory processing further advances the field of robotic locomotion [11], enabling more dynamic approaches and greater adaptability to diverse scenarios. DRL offers a promising framework for handling stochastic environments, and adapting to challenging conditions [12]–[24].

In this study, we comparatively examine learning-based control strategies, a biologically inspired gait design method, and their hybrid framework on a salamander-inspired robot (Fig. 1). Specifically, we evaluate two distinct robot configurations: one employing a fixed spinal joint and another featuring an active spinal joint. Initially, the salamander robot utilizes a predefined footfall pattern based on the Hildebrand gait method [25], with a fixed spinal joint facilitating linear locomotion. Subsequently, the robot retains this footfall pattern but incorporates an active spinal joint, which enables dynamic body undulation achieved through a DRL method. To enable precise control and facilitate

¹Department of Computer Science, University of Southern California, Los Angeles, CA 90089 ²Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 ³Department of Electrical and Computer Engineering, University of Illinois Chicago, Chicago, IL 60607,*These authors contributed equally. corresponding author: yozkanay@nd.edu

effective learning, we developed a digital twin of the physical robot integrated within the MuJoCo simulation environment [26], evaluating the system across various terrain scenarios.

Prior work has explored several complementary directions for legged and salamander-inspired locomotion. Biologically detailed salamander models have studied spinal-cord mechanisms, sensory feedback, and spontaneous gait transitions [1], [3], [5]. In quadruped robotics, hybrid CPG-RL methods learn to modulate oscillator parameters for robust and omnidirectional locomotion, including successful sim-to-real transfer on conventional quadrupeds [27]. More recently, phase-based learning methods for salamander-inspired robots have targeted diverse and omnidirectional gait repertoires by allowing different body parts to evolve through learned phase variables [28]. In contrast, our goal is not to learn a full multi-gait or omnidirectional controller. We instead ask a narrower design question: whether a minimal active spinal joint can improve crawling performance on a small salamander-inspired platform when the limb gait is kept deliberately structured. This choice reduces the learning problem and facilitates hardware deployment.

The main contributions of our work are as follows:

- We present a systematic study of crawling locomotion in a salamander-inspired quadruped, expanding legged-robotics research beyond the more commonly studied dog-like morphologies.
- We introduce a hybrid control framework that combines a biologically inspired Hildebrand footfall pattern with deep reinforcement learning for spinal actuation, enabling a balance between structured coordination and adaptive behavior.
- We perform a comprehensive empirical evaluation across fixed-spine and active-spine robot configurations, including simulation experiments and sim-to-real transfer on hardware, to assess the effectiveness and practical feasibility of the proposed approach.

II. MATERIALS AND METHODS

A. Robot Design & Actuation

The robot design is inspired by the biomechanics of the salamander after conducting a comprehensive review of amphibious and sprawling locomotion [29], [30]. To strike a balance between biological fidelity and mechanical simplicity, we developed a simplified quadrupedal model. Each of the four legs has 2 degrees of freedom (DoF), allowing for controlled motion in the fore-aft (protraction/retraction) and up-down (elevation/depression) planes. To replicate the body undulation characteristic of a sprawling gait, a single 1-DoF spinal joint was integrated into the robot's trunk. The body joint facilitates lateral bending of the front and rear

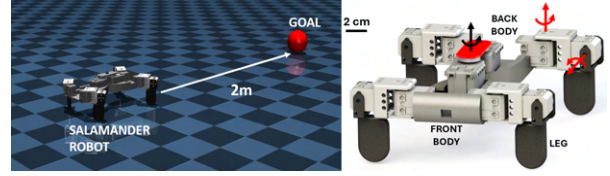


Fig. 2. **Simulation Environment and Robot Actuation.** (Left) The MuJoCo simulation environment used for evaluating the robot's locomotion. The red sphere indicates the target goal. (Right) CAD model of the robot detailing its kinematic structure. The red arrows represent the two rotational axes per leg, and the black arrow indicates the single spinal joint that enables lateral bending for sprawling locomotion.

body segments in the horizontal plane. The use of a 1-DoF spine was chosen as a minimal morphology that captures lateral body bending while keeping the robot mechanically simple, easier to model, and feasible for hardware deployment. This design does not attempt to reproduce the full distributed flexibility of a biological salamander spine. Instead, it provides a controlled way to study whether even a single active lateral-bending joint can improve crawling performance when combined with structured limb coordination.

The salamander robot is designed using SolidWorks (Fig-2), and mechanical components, including the legs, shoulders, and front and rear body segments, are 3D printed using ABS filament with a Stratasys F170 3D printer. The leg movements are powered by Dynamixel XL-320 servo motors, while a Dynamixel AX-12 servo actuates the spinal joint. These servos provide real-time feedback on both torque and position. A U2D2 Power Hub Board handles power distribution, using TTL and RS-485 communication to connect the servo motors to the controller. This design, visually represented in Fig-2, effectively captures the essential biomechanics of salamander movement.

B. Open-Loop Hildebrand Gait Design Method

In this study, we utilize the Hildebrand gait analysis method, a well-established classification system for characterizing quadrupedal locomotion based on the timing and sequencing of footfalls [31], [32]. This method has been widely adopted in both biomechanical studies of animals and the control of legged robots [7], [10]. Hildebrand's framework defines symmetrical gaits using two key kinematic parameters: the duty factor and the leg phase shift. The duty factor (β) is the percentage of a full gait cycle during which a single foot is in contact with the ground. Gaits with a duty factor greater than 0.5 are classified as walks, as at least two feet are always on the ground, ensuring static stability. The leg phase shift (ϕ) is the temporal offset, expressed as a percentage of the gait cycle, between the footfalls of ipsilateral (same-side) limbs. For example, a phase shift of 0.5 (or 50%)

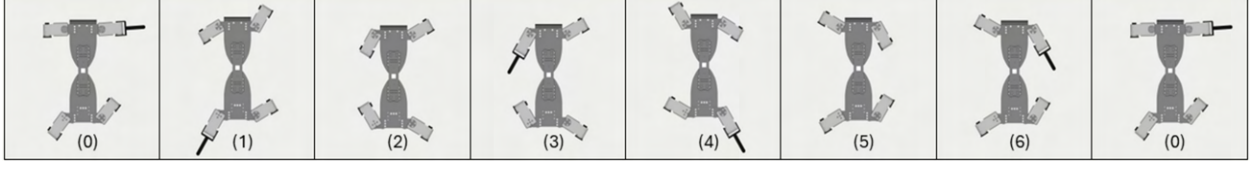


Fig. 3. **Example biological gait-cycle.** Joint positions during one walking cycle following the Hildebrand-style gait for both the passive and active spinal joint scenarios. Each leg spends 25% of the cycle in the air and 75% on the ground.

represents a trot-like gait where diagonal pairs of legs move in unison.

In this work, we employed a specific walking gait characterized by a duty factor of 0.75, meaning each leg spends 75% of the cycle in the stance phase and 25% in the swing phase. This high duty factor ensures continuous ground contact and inherent stability. The leg phase shift was set to 0.25 (25%), corresponding to a lateral-sequence gait, where the footfall sequence on one side of the body is hind-fore, then the same sequence on the other side. This gait pattern is particularly common in sprawling animals like salamanders and offers a balance between stability and maneuverability. Figure 3 illustrates the commanded joint angles for each leg throughout a single gait cycle. This cyclic pattern is continuously iterated as the robot navigates toward its target.

C. Deep Reinforcement Learning

Reinforcement learning (RL) provides a general framework for solving sequential decision-making problems, which underpins much of robot learning [33]–[38]. In line with prior work on quadruped robots, we train our salamander-inspired robot using both Proximal Policy Optimization (PPO) [39] and Soft Actor-Critic (SAC) [36]. Since SAC consistently outperforms PPO in our experiments, we report results using SAC. Soft Actor-Critic is an off-policy RL algorithm that extends actor-critic methods by incorporating entropy regularization to incentivize exploration:

$$\pi^* = \arg \max_{\pi} \mathbb{E}_{\tau \sim \pi} [\sum_{t=0}^{\infty} \gamma^t (R(s_t, a_t, s_{t+1}) + \alpha H(\pi(\cdot | s_t)))] \quad (1)$$

where

$$H(X) = - \sum_{x \in \mathcal{X}} P(x) \log P(x) \quad (2)$$

represents the entropy term. Here, $\alpha > 0$ is a trade-off parameter that balances the reward function and the entropy term, and hence exploration and exploitation [36], [40]–[42]. The SAC algorithm is capable of handling high-dimensional observation and continuous action spaces while maintaining stability during learning.

Fig-4 (left) presents the observation space for both the 8-joint (no spinal joint) and 9-joint robot configurations,

while the action space consists of joint angles (in radians) in our setup. The reward function is inspired by MuJoCo’s ant environment and previous work done on quadruped robots [43], [44] to encourage efficient moves towards the goal while maintaining stability. It consists of multiple components:

$$R(s, a) = w_1 \cdot \Delta x + w_2 \cdot \Delta d + w_3 \cdot \Delta y + w_4 \cdot C + H \quad (3)$$

where Δx represents the change in the x-direction to incentive forward movement, Δd indicates the change in distance to the goal, Δy is the change in the y-direction to incentivize the stability and ensure smooth movement, C is control cost term to punish the robot for taking excessively large actions, and H represents the constant healthy reward (to prevent the robot from jumping or flipping), and $w_1 = 5, w_2 = 150, w_3 = 5, w_4 = 0.1$ are the corresponding weights. The reward weights are selected through preliminary tuning to prioritize reaching the target while discouraging excessive lateral drift and large actions.

The hyperparameters used for training the reinforcement learning model are listed in Table I. These include the batch size, learning rate, number of steps collected before learning begins, discount factor (which balances short-term vs. long-term rewards), entropy coefficient (which controls exploration-exploitation trade-off), and soft update coefficient (which governs the update rate of the target network). All parameters are tuned to optimize learning stability and performance.

The two components of the controller play different roles. The Hildebrand gait specifies the cyclic limb timing and ensures that the robot maintains a statically stable walking pattern with frequent ground contact. SAC is then used only for the spinal joint, where the policy can exploit sensory feedback to adjust lateral bending and improve forward progression. This separation reduces the action dimension from all limb and spine joints to the spinal action alone in the hybrid setting, while preserving the ability to learn morphology-specific body motion.

III. EXPERIMENTAL SETUP & RESULTS

The simulation environment is built using MuJoCo [26]. During simulation, the robot employs Hildebrand

STATE VARIABLES	8-JOINTS	9-JOINTS
X/Y/Z COORDINATES OF TORSO	3	3+3
X/Y/Z/W ORIENTATIONS OF TORSO	4	4+4
LIMB JOINT ANGLES	2x4	2x4
X/Y/Z VELOCITIES OF TORSO	3	3+3
X/Y/Z ANGULAR VELOCITIES OF TORSO	3	3+3
ANGULAR VELOCITIES OF LEG JOINTS	2x4	2x4
SPINAL JOINT ANGLE	0	1
SPINAL ANGULAR VELOCITY	0	1
DISTANCE TO GOAL	3	3
ANGLE TO GOAL	1	1
FRICTIONS (X/Y/Z) ON LEGS	3x4	3x4
TOTAL	45	60

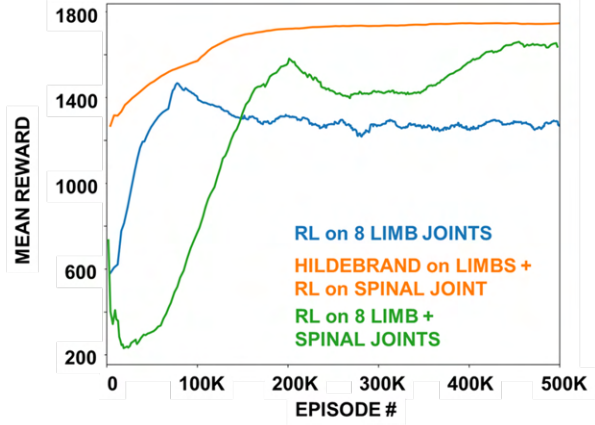


Fig. 4. **State variables and Results.** (Left) Observation space for the 8-joint and 9-joint robot configurations. (Right) RL Learning Curves.

gaits or DRL actions to locomote toward a goal, a red ball positioned 2 meters away from the initial position (Fig.2).

The gaits exhibit cyclic behavior, i.e. repeats at a fixed period. However, this periodicity does not disrupt forward motion, and advance the robot smoothly toward the target. At each discrete time point, joint angles (in radians) are prescribed to the robot’s actuators, and forward kinematics calculations update its position. The simulation concludes once the robot reaches the goal after a predetermined number of iterations.

In this section, we assess the efficiency of our hybrid RL–Hildebrand framework and try to answer two core questions:

- 1) How does end-to-end RL compare to classical Hildebrand gaits in terms of goal accuracy and stability?
- 2) What is the impact of an active spinal joint versus a fixed spine on overall crawling locomotion performance?

To this end, we evaluate all methods on a standardized target-directed navigation task and conducted a series of comparative studies, testing four distinct locomotion policies: (1) a traditional Hildebrand-only gait, (2) an unconstrained reinforcement learning (RL) policy, (3) an RL policy with torque constraint, and (4) our proposed hybrid RL–Hildebrand framework with an actively controlled spinal joint. The objective was to systematically assess how spinal flexibility and the integration of biologically inspired priors influence locomotion robustness and performance.

Experiments were conducted over 100 simulated episodes, and the average results are reported. The comparison of the models is performed in the following way by using three key metrics: (1) **MDB**: the minimum distance to the goal over the full trajectory. (2) **DY (Lateral Deviation)**: the maximum displacement along

TABLE I
HYPERPARAMETERS FOR 8-JOINTS AND 9-JOINTS SETTINGS.

Hyperparameters	8-joints	9-joints
Batch size	128	128
Learning starts	5000	5000
Discount factor (γ)	0.958	0.972
Learning rate	0.0007	0.0053
Entropy coefficient (α)	0.0815	0.04
Soft update parameter (τ)	0.006	0.016

the y-axis over the full trajectory. (3) **ATB**: the average timesteps to reach the goal.

An episode is marked as successful if

$$MDB \leq 0.15 \quad \text{and} \quad DY \leq 0.05. \quad (4)$$

For successful episodes, ATB quantifies traversal efficiency. While a lower ATB indicates faster navigation, we emphasize that excessively rapid trajectories may be infeasible on the physical salamander-inspired robot, and hence is not necessarily better in practice.

A. SAC vs Hildebrand Gaits for 8-joint Robot Configuration

Our experimental evaluation first aimed to address (1) How does an end-to-end reinforcement learning (RL) policy compare to classical Hildebrand gaits in terms of goal accuracy, speed, and stability? To investigate this, we conducted a comparative study of the 8-joint robot, which lacks an active spinal joint, using three distinct control strategies: a traditional Hildebrand gait, an unconstrained RL policy, and a torque-limited RL policy (denoted as RL*, during inference, the actions/angles obtained from the RL model are clipped by applying torque limits to the motors).

TABLE II

PERFORMANCE COMPARISON OF DIFFERENT APPROACHES FOR THE 8-JOINT CONFIGURATION. VALUES ARE PRESENTED AS MEAN $\pm$ STANDARD DEVIATION OVER 5 SEEDS. METRICS: MDB (MINIMUM DISTANCE TO THE BALL), DY (DEVIATION FROM THE Y-AXIS), AND ATB (AVERAGE TIMESTEPS TO GOAL).

8-joints Version	MDB	DY	ATB
Hildebrand	0.10	0.03	6103
RL	0.04 ± 0.002	0.02 ± 0.001	104 ± 1
RL*	0.13 ± 0.039	0.02 ± 0.001	334 ± 17

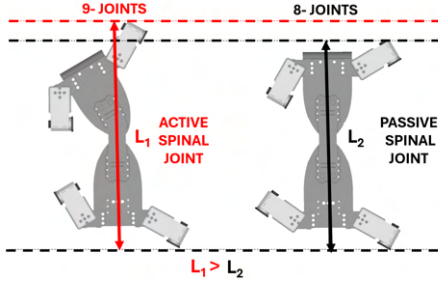


Fig. 5. **Reachability with and without an Active Spinal Joint.** The figure compares the forward reachability of the robot with a flexible spine (9-joints) and a rigid spine (8-joints). As shown, the active spinal joint enables the robot to extend its reach farther forward ($L_1 > L_2$) for the same leg joint angles.

Table II illustrates the performance of the robot under these three conditions. The Hildebrand gait, while producing robust and easily executed motions, yielded relatively slow, walking-like locomotion. In contrast, an unconstrained RL agent learns high-speed, cheetah-like gaits that exceed the salamander robot’s physical capabilities. While this behavior was highly effective in the idealized simulation, it is not physically realizable on our hardware. To bridge this gap between simulation and reality, we introduced actuator force limits ($[-0.39, 0.39]$) to the RL policy, creating the RL^* strategy. This simple constraint successfully curbed excessive velocities, resulting in a policy that was both significantly faster than the Hildebrand baseline and more stable than the pure RL policy.

B. Fixed vs Active Spinal Joint Robot Configurations

We address the second core question: what is the impact of an active spinal joint versus a fixed spine on overall locomotion performance? As shown in Figure 5, an active spinal joint significantly expands the robot’s kinematic workspace. For identical leg joint angles, the 9-joint configuration with an active spine achieves a greater forward reach than the 8-joint configuration with a fixed spine. This suggests that adding an active spinal joint increases the reachable body configurations and can improve forward progression for the tested crawling task.

TABLE III

PERFORMANCE COMPARISON OF DIFFERENT APPROACHES FOR THE 9-JOINT CONFIGURATION. VALUES ARE PRESENTED AS MEAN $\pm$ STANDARD DEVIATION OVER 5 SEEDS. METRICS: MDB (MINIMUM DISTANCE TO THE BALL), DY (DEVIATION FROM THE Y-AXIS), AND ATB (AVERAGE TIMESTEPS TO GOAL).

Method	MDB	DY	ATB
Hildebrand + RL	0.05 ± 0.001	0.03 ± 0.0008	1518 ± 6
RL	0.07 ± 0.008	0.04 ± 0.001	196 ± 2
RL*	0.10 ± 0.004	0.05 ± 0.002	1931 ± 1

Table III compares the performance of the 9-joint robot across three distinct control strategies: a hybrid model (Hildebrand gaits for legs + RL for the spinal joint), a pure end-to-end RL policy for all joints, and a torque-constrained RL policy (RL^*) for all joints.

Unlike the 8-joint configuration, where imposing torque limits on a pure RL policy led to improved locomotion, directly constraining the end-to-end RL policy on all 9 joints did not yield a stable or efficient gait. Instead, we found that the most effective strategy was our hybrid approach, which uses Hildebrand gaits to control the rhythmic leg movements and an RL policy to adaptively control the spinal joint. This produced a more efficient and physically feasible locomotion pattern for our salamander-inspired robot. These results suggest that, for our robot and crawling task, combining a structured limb gait with learned spinal control can produce more physically feasible locomotion than learning all joints end-to-end.

C. Locomotion on Rough Terrain

To generate the rough terrain used in our simulations, we procedurally synthesized a heightmap by combining multiple octaves of Gaussian-filtered random noise. Each octave was normalized and scaled according to persistence and lacunarity parameters, allowing us to capture both large-scale elevation changes and fine-grained irregularities. We evaluated the performance of several strategies on two distinct terrain levels, categorized as

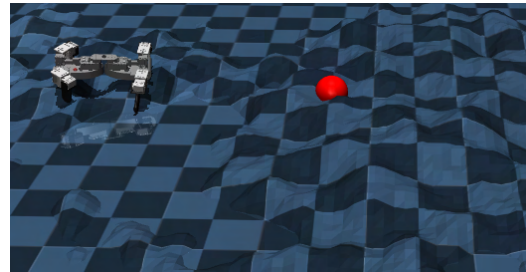


Fig. 6. Simulation of the salamander robot on rough terrain in MuJoCo Physics Simulator.

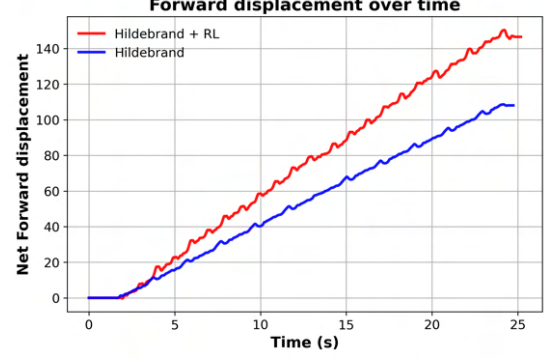
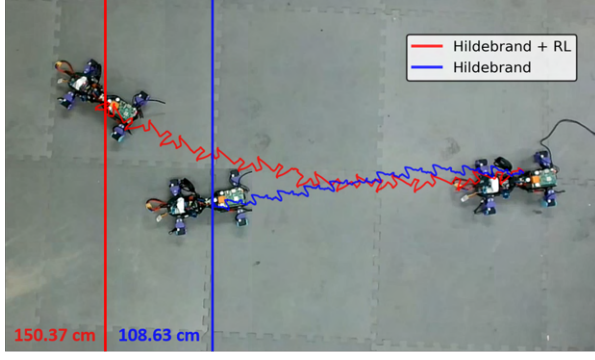


Fig. 7. Comparison of real-world performance: (Left) Trajectories from motion capture. Hybrid Hildebrand+RL (red): 150.4 cm; Hildebrand-only (blue): 108.6 cm. (Right) Forward displacement over time for the same trial, showing a 38% speed increase.

easy and hard (Table IV, Fig. 6). Similarly, we compared three control strategies: an 8-joint RL model, a 9-joint RL model (with an added joint for spinal control), and a combined Hildebrand + RL approach. For both the 8-joint RL and 9-joint RL models, we also tested a torque-limited version denoted with an asterisk (*). Performance was quantified using two metrics, MDB and ATB, described in Section 3. The simulation results on rough terrain support the effectiveness of our hybrid approach. Among all methods, only the 8-joint RL baseline and our hybrid approach consistently reach closer to the target ball.

TABLE IV
PERFORMANCE ON TWO LEVELS OF TERRAIN.

	Easy Level Terrain		Hard Level Terrain	
	MDB	ATB	MDB	ATB
8-Joint RL	0.40	540	0.57	350
8-Joint RL*	0.77	449	1.35	217
9-Joint RL	0.31	307	0.66	946
9-Joint RL*	1.41	1002	2.31	1002
Hildebrand + RL	0.49	1674	0.29	3358



Fig. 8. Qualitative evaluation of the robot on uneven terrain. The hybrid Hildebrand+RL strategy enabled successful traversal on (left) rocks (sizes between 2-4 cm) and (right) dry sand.

D. Sim-to-Real Gap: Observations

While simulation experiments provided valuable insights regarding locomotion strategies, physical implementation on the salamander-inspired robot revealed limitations that highlight the sim-to-real gap. For both 9-joint (active spine) and 8-joint (rigid spine) configurations, reinforcement learning policies trained in simulation, whether unconstrained (RL) or torque-limited (RL*), proved infeasible on hardware. As demonstrated in the Supplementary Information (SI) Movie (0:13–0:26), these policies generated high-frequency, cheetah-like gaits that exceeded the torque and velocity capabilities of the physical robot’s actuators. Even torque-limited policies produced oscillatory, unstable motions, confirming that behaviors optimized in idealized environments did not generalize to the physical platform.

Thus, only two locomotion plans proved appropriate for real-world test: Hildebrand gait in the 8-joint configuration and Hildebrand+RL hybrid framework with Hildebrand gait controlling the legs and RL policy controlling the spinal joint in the 9-joint configuration. This limitation defines the importance of incorporating biologically inspired priors in control policies to make them physically realizable.

Real-world trials showed that the hybrid Hildebrand+RL configuration achieved an average of 38% faster traversal speed relative to the Hildebrand-only baseline, while maintaining stable, goal-directed locomotion. This improvement was primarily reflected in the increased forward displacement over the same time interval, highlighting the contribution of active spinal control to overall locomotor speed.

As shown in Figure 7(a), the hybrid configuration achieved a forward displacement of 150.4 ± 2.1 cm in the same 24-second interval where the Hildebrand-only gait reached 108.6 ± 5.4 cm. Figure 7(b) further quantifies this difference, showing consistently greater

forward progress over time. These plots correspond to a single representative trial, included for illustrative purposes, while the averaged results across repeated trials are reported in Table V. Lateral deviations observed in both cases stem from the sim-to-real gap, including differences in weight distribution between the robot and its digital twin, unmodeled surface interactions, and reliance on open-loop actuation.

TABLE V

COMPARISON OF REAL-WORLD LOCOMOTION PERFORMANCE. VALUES SHOW MEAN FORWARD DISPLACEMENT AND AVERAGE SPEED AFTER 24 SECONDS (AVERAGED OVER REPEATED TRIALS).

Method	Forward displacement (cm)	Speed (cm/s)
Hildebrand (8-joint)	108.6 $\pm$ 5.4	4.54
Hybrid Hildebrand+RL (9-joint)	150.4 $\pm$ 2.1	6.27

In addition to flat-ground trials, we conducted preliminary qualitative tests on uneven terrain, including rocks (2-4 cm) and dry sand. In both cases, the hybrid Hildebrand+RL strategy enabled successful traversal, as illustrated in Figure 8 (see also SI Movie, 0:48–0:56). These results indicate that active spinal control contributes to robustness by redistributing its weight to maintain stability and traction, beyond controlled laboratory conditions. A more systematic quantitative evaluation of locomotion across varied terrains will be pursued as part of future work.

These findings show that, in our current implementation, RL policies trained in simulation did not directly transfer to the physical salamander-inspired robot. We attribute this to a combination of simplified actuator modeling, unmodeled servo dynamics, ground-contact modeling errors, and the sprawled morphology of the robot. This should not be interpreted as evidence that RL locomotion policies are generally non-transferable. Rather, it indicates that successful transfer of fully learned policies likely requires additional actuator identification, domain randomization, or hardware-aware constraints, whereas our hybrid controller reduces the transfer burden by preserving a physically interpretable limb gait.

In contrast, our results demonstrate that a carefully structured hybrid approach can successfully mitigate a significant portion of the sim-to-real gap by explicitly accounting for the unique kinematic and dynamic properties of the sprawled architecture. At the same time, they highlight unresolved challenges that motivate our future research directions, including more systematic terrain evaluation and biologically inspired control priors.

E. Future Work and Conclusion

A primary avenue for future research lies in refining the reward function, a critical component of reinforcement learning that significantly influences the learned behavior. The hand-crafted reward schemes commonly

used in quadrupedal robotics, which often penalize energy consumption while rewarding forward velocity, can inadvertently lead to undesirable or unsafe behaviors. To address this, we propose leveraging formal specifications through Signal Temporal Logic (STL) to systematically and more safely shape the reward function [45], [46]. This STL-based approach provides a principled and robust alternative to heuristic reward design, offering greater sample efficiency and a more formal guarantee of desired robot behaviors.

Overall, our results suggest that a minimal active spinal joint can improve forward crawling performance on a salamander-inspired robot when combined with a structured limb gait. The hybrid Hildebrand–RL controller provides a practical compromise: it reduces the dimensionality of the learning problem and improves hardware feasibility, but at the cost of limiting gait flexibility. More broadly, this study highlights the potential of combining biologically inspired, model-based locomotion priors with the adaptability of data-driven learning. Our approach is therefore best suited for studying the contribution of spinal actuation in constrained crawling tasks, while future work should extend the framework toward adaptive gait parameters, richer spine-control ablations and broader terrain evaluation.

REFERENCES

- [1] L. T. Phan, Y. H. Lee, Y. H. Lee, H. Lee, H. Kang, and H. R. Choi, "Study on effects of spinal joint for running quadruped robots," *Intelligent Service Robotics*, vol. 13, no. 1, pp. 29–46, 2020.
- [2] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science robotics*, vol. 5, no. 47, p. eabc5986, 2020.
- [3] A. Pazzaglia, A. Bicanski, A. Ferrario, J. Arreguit, D. Ryczko, and A. Ijspeert, "Balancing central control and sensory feedback produces adaptable and robust locomotor patterns in a spiking, neuromechanical model of the salamander spinal cord," *PLOS Computational Biology*, vol. 21, no. 1, p. e1012101, 2025.
- [4] S. Suzuki, T. Kano, A. J. Ijspeert, and A. Ishiguro, "Spontaneous gait transitions of sprawling quadruped locomotion by sensory-driven body–limb coordination mechanisms," *Frontiers in Neurobotics*, vol. 15, p. 645731, 2021.
- [5] A. J. Ijspeert, A. Crespi, D. Ryczko, and J.-M. Cabelguen, "From swimming to walking with a salamander robot driven by a spinal cord model," *science*, vol. 315, no. 5817, pp. 1416–1420, 2007.
- [6] B. Chong, Y. O. Aydin, C. Gong, G. Sartoretti, Y. Wu, J. M. Rieser, H. Xing, J. W. Rankin, K. B. Michel, A. G. Nicieza, *et al.*, "Coordination of back bending and leg movements for quadrupedal locomotion," in *Robotics: Science and Systems*, vol. 20. Pittsburgh, PA, 2018.
- [7] B. Chong, Y. Ozkan Aydin, G. Sartoretti, J. M. Rieser, C. Gong, H. Xing, H. Choset, and D. I. Goldman, "A hierarchical geometric framework to design locomotive gaits for highly articulated robots," in *Robotics: science and systems*, 2019.
- [8] Y. Ozkan Aydin, B. Chong, C. Gong, J. M. Rieser, J. W. Rankin, K. Michel, A. G. Nicieza, J. Hutchinson, H. Choset, and D. I. Goldman, "Geometric mechanics applied to tetrapod locomotion on granular media," in *Conference on biomimetic and biohybrid systems*. Springer, 2017, pp. 595–603.
- [9] A. J. Ijspeert, "Amphibious and sprawling locomotion: From biology to robotics and back," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, no. 1, pp. 173–193, 2020.

- [10] B. Chong, Y. O. Aydin, C. Gong, G. Sartoretti, Y. Wu, J. M. Rieser, H. Xing, P. E. Schiebel, J. W. Rankin, K. B. Michel, et al., "Coordination of lateral body bending and leg movements for sprawled posture quadrupedal locomotion," *The International Journal of Robotics Research*, vol. 40, no. 4-5, pp. 747-763, 2021.
- [11] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238-1274, 2013. [Online]. Available: <https://doi.org/10.1177/0278364913495721>
- [12] H. Benbrahim and J. A. Franklin, "Biped dynamic walking using reinforcement learning," *Robotics and Autonomous systems*, vol. 22, no. 3-4, pp. 283-302, 1997.
- [13] R. Tedrake, T. W. Zhang, and H. S. Seung, "Stochastic policy gradient reinforcement learning on a simple 3d biped," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, vol. 3. IEEE, 2004, pp. 2849-2854.
- [14] X. B. Peng, G. Berseth, and M. Van de Panne, "Dynamic terrain traversal skills using reinforcement learning," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, pp. 1-11, 2015.
- [15] —, "Terrain-adaptive locomotion skills using deep reinforcement learning," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1-12, 2016.
- [16] X. B. Peng, G. Berseth, K. Yin, and M. Van De Panne, "Deeploco: Dynamic locomotion skills using hierarchical deep reinforcement learning," *Acm transactions on graphics (tog)*, vol. 36, no. 4, pp. 1-13, 2017.
- [17] N. Heess, D. Tb, S. Sriram, J. Lemmon, J. Merel, G. Wayne, Y. Tassa, T. Erez, Z. Wang, S. Eslami, et al., "Emergence of locomotion behaviours in rich environments," *arXiv preprint arXiv:1707.02286*, 2017.
- [18] Z. Fu, A. Kumar, J. Malik, and D. Pathak, "Minimizing energy consumption leads to the emergence of gaits in legged robots," *arXiv preprint arXiv:2111.01674*, 2021.
- [19] G. Bellegarda, Y. Chen, Z. Liu, and Q. Nguyen, "Robust high-speed running for quadruped robots via deep reinforcement learning," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 10 364-10 370.
- [20] Y. Ji, Z. Li, Y. Sun, X. B. Peng, S. Levine, G. Berseth, and K. Sreenath, "Hierarchical reinforcement learning for precise soccer shooting skills using a quadrupedal robot," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 1479-1486.
- [21] M. Aractingi, P.-A. Léziart, T. Flayols, J. Perez, T. Silander, and P. Souères, "Controlling the solo12 quadruped robot with deep reinforcement learning," *scientific Reports*, vol. 13, no. 1, p. 11945, 2023.
- [22] L. Han, Q. Zhu, J. Sheng, C. Zhang, T. Li, Y. Zhang, H. Zhang, Y. Liu, C. Zhou, R. Zhao, et al., "Lifelike agility and play in quadrupedal robots using reinforcement learning and generative pre-trained models," *Nature Machine Intelligence*, vol. 6, no. 7, pp. 787-798, 2024.
- [23] D. Hoeller, N. Rudin, D. Sako, and M. Hutter, "Anymal parkour: Learning agile navigation for quadrupedal robots," *Science Robotics*, vol. 9, no. 88, p. eadi7566, 2024.
- [24] G. Bellegarda, C. Nguyen, and Q. Nguyen, "Robust quadruped jumping via deep reinforcement learning," *Robotics and Autonomous Systems*, vol. 182, p. 104799, 2024.
- [25] M. Hildebrand, "Chapter 3. walking and running," in *Functional vertebrate morphology*. Harvard University Press, 1985, pp. 38-57.
- [26] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 5026-5033.
- [27] G. Bellegarda and A. Ijspeert, "Cpg-rl: Learning central pattern generators for quadruped locomotion," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12 547-12 554, 2022.
- [28] Z. Liu, Y. Liu, Y. Fang, and X. Guo, "Learning omnidirectional locomotion for a salamander-like quadruped robot," *arXiv preprint arXiv:2511.08299*, 2025.
- [29] A. J. Ijspeert, "Amphibious and sprawling locomotion: From biology to robotics and back," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, no. 1, pp. 173-193, 2020. [Online]. Available: <https://doi.org/10.1146/annurev-control-091919-095731>
- [30] Z. Bing, A. Rohregger, F. Walter, Y. Huang, P. Lucas, F. O. Morin, K. Huang, and A. Knoll, "Lateral flexion of a compliant spine improves motor performance in a bioinspired mouse robot," *Science Robotics*, vol. 8, no. 85, p. eadg7165, 2023. [Online]. Available: <https://www.science.org/doi/abs/10.1126/scirobotics.adg7165>
- [31] M. Hildebrand, "Analysis of Asymmetrical Gaits," *Journal of Mammalogy*, vol. 58, no. 2, pp. 131-156, 05 1977. [Online]. Available: <https://doi.org/10.2307/1379571>
- [32] —, "Symmetrical gaits of horses: Gaits can be expressed numerically and analyzed graphically to reveal their nature and relationships," *Science*, vol. 150, no. 3697, pp. 701-708, 1965.
- [33] R. S. Sutton, A. G. Barto, et al., *Reinforcement learning: An introduction*. MIT press Cambridge, 1998, vol. 1, no. 1.
- [34] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *International conference on machine learning*. PMLR, 2015, pp. 1889-1897.
- [35] S. Gu, E. Holly, T. Lillicrap, and S. Levine, "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 3389-3396.
- [36] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," Aug. 2018, arXiv:1801.01290 [cs]. [Online]. Available: <http://arxiv.org/abs/1801.01290>
- [37] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International conference on machine learning*. PMLR, 2018, pp. 1587-1596.
- [38] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for versatile, dynamic, and robust bipedal locomotion control," *The International Journal of Robotics Research*, vol. 44, no. 5, pp. 840-888, 2025.
- [39] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *ArXiv*, vol. abs/1707.06347, 2017. [Online]. Available: <https://api.semanticscholar.org/CorpusID:28695052>
- [40] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *International conference on machine learning*. PmlR, 2016, pp. 1928-1937.
- [41] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.
- [42] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-dimensional continuous control using generalized advantage estimation," *arXiv preprint arXiv:1506.02438*, 2015.
- [43] R. Boney, J. Sainio, M. Kaivola, A. Solin, and J. Kannala, "RealAnt: An Open-Source Low-Cost Quadruped for Education and Research in Real-World Reinforcement Learning," June 2022, arXiv:2011.03085 [cs]. [Online]. Available: <http://arxiv.org/abs/2011.03085>
- [44] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke, "Sim-to-Real: Learning Agile Locomotion For Quadruped Robots," May 2018, arXiv:1804.10332 [cs]. [Online]. Available: <http://arxiv.org/abs/1804.10332>
- [45] A. Balakrishnan and J. V. Deshmukh, "Structured reward shaping using signal temporal logic specifications," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 3481-3486.
- [46] S. Kulgod, W. Chen, J. Huang, Y. Zhao, and N. Atanasov, "Temporal logic guided locomotion planning and control in cluttered environments," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 5425-5432.